Diane Vogt-O'Connor

Is the Record of the 20th Century at Risk?

At the end of the 20th century the world's archives are facing challenges of a scale that were unimaginable only a decade ago. First, the size of the human record has grown at a geometric progression beginning with the development of the photocopy machine and desktop publishing and greatly enhanced by electronic communications. The National Archives and Records Administration is now annually accepting 10 times more electronic records from the Treasury Department in email alone than it received from the entire federal government in the previous 25 years according to a recent article by Archivist John Carlin.[1]

Not only are the old creators of records writing and producing more documentation, but new voices—new organizations, new individuals, and new groups—are being empowered by these media to produce documentation reflecting their views. Every eight minutes, more new information is added to the Internet alone than is currently held by the United States National Archives and Records Administration.[2]

This wealth of digital data provides us with a chance to learn more about humans as a diverse, often frivolous, and endlessly inventive species than any previous media since the printing press. We can explore our world and its cultural and natural resources without leaving home. Yet this elegant and playful electronic record is endangered. Unlike our photographic, paper, and motion picture film, which gradually decline into full loss over a period of centuries, digital media self-destruct in decades with little warning. Some vanish much more speedily—the average life expectancy of a web page is roughly 70 days according to Brewster Kahle of the Internet Archive.[3] **We may paradoxically be both the best documented era in history and the least understood, as much of our documentation will be lost.**[4]

While we have the first telephone and telegraph messages still, the first email message, chat group session, and web site have already been lost. The predicted pace of electronic information loss is accelerating. Not only do we have to contend with the fragility of digital media; even more worrisome is the speedy obsolescence of the software and hardware that makes the files usable.[5] These market-driven systems change rapidly—roughly every 18 months—and are often not able to play earlier files, leaving historic data orphaned and inaccessible. Since this data is our cumulative memory as a species, the situation is dire.

The market incentive for software and hardware manufacturers to step in and solve this problem for society doesn't yet exist. The vendors profit from system obsolescence by selling corporations, groups and individuals new or upgraded software and hardware and new formats of old content over and over again. How many different versions of your favorite music do you have (records, tapes, digital tapes, CD-ROMs, DVDs)? Few groups or individuals have budgeted to keep moving old content to new formats endlessly; therefore, we lose significant portions of our heritage of data, information, and knowledge daily.

Why Do Archivists Fear a New Dark Ages?

Many electronic preservation specialists believe that a significant portion of late-20th-century data, information, and knowledge will be lost permanently. In a letter to the editor of the *Washington Post* and in the newsletter *CLIR Issues,* Deanna Marcum of the Council of Library and Information Resources has predicted a 10-year loss of digital records.[6] Danny Hillis of Walt Disney predicts "A 'digital gap' will span from the beginning of the wide-spread use of the computer until the time we eventually solve this problem."[7] Peter Lyman of the University of California at Berkeley asks "Are digital signals destined to be a kind of oral culture, living only as long as they are remembered and repeated?" Our era may in effect become a new dark age about which most of the core information, knowledge, and data will be lost—except for that printed to paper or continuously migrated to newer software platforms.[8]

Conservators point to the significant losses already experienced with such 20th-century media as color photographs, cellulose nitrate motion picture film, audiotapes, videotapes, and high-lignin wood pulp paper records. When combined with the predicted loss of digital files the effect may be a general loss of contemporary memory, perhaps the greatest such loss since the 1400s.[9]

According to many experts, to prevent the emergence of such loss we must develop:

- **a universal preservation format** that ideally is long-lived, compressed, but still eye legible
- **a universal translator** that is able to move old files to the new formats constantly being developed by our market-driven economy [10]
- **certified repositories** that are adequately funded with trained and imaginative staff who are well equipped to ensure the survival of our new major documentation and communications media—particularly digital data. Currently NARA has only 2 million dollars annually to manage electronic records.[11]

When this is contrasted with the 40+ million dollars currently being spent by the National Digital Library to move stable paper records into short-lived, but accessible digital formats, it rapidly becomes clear that preservation of knowledge has not been given equal priority to access in this country.

To keep electronic files, they must be migrated (moved to the next generation of hardware and software) and refreshed (copied to new and more durable media as digital media itself is fragile and short-lived and given a new tape wind to limit stresses) every 18 months or so.[12]

Costs of Managing Digital Data Over Time

Electronic records project experts have estimated that digital records are roughly 10-16 times more expensive to manage over time than paper records. National Digital Library Ameritech grant experience indicates that the cost of digitizing an item is only one third of the start up cost of digital work, with two thirds being the cost of cataloging, metadata, and quality control.[13]

Archives are not funded to a level that empowers them to deal with the long-term management of the growing quantities of digital data, nor are most archivists trained to work with these media. Yet ignoring the problem is not an option. In recent court cases, archivists at the National Archives have been held responsible for managing electronic versions of federal documents effectively, regardless of institutional abilities and funding.[14]

These demands and legal requirements for super-archives to save an ever greater and more diverse record are coming at a time when:

- organizations and governments are reorganizing and downsizing
- archival budgets are flat or in decline
- archival descriptive standards are in flux[15]
- archival staff must retrain to learn the new standards
- legislation affecting archives is changing

What Legal Challenges Affect Archives?

Legal standards are also in flux. Archivists are currently facing some of the most stringent legal challenges ever to their right to provide fair use access to materials whose copyrights are held by others. Recent legislation has extended the duration of copyright protection by 20 more years. Archivists' traditional role has been to provide access under the legal concept of "fair use," which allows access for scholarship, parody, education, and news reporting purposes. Recent rulings by the courts and developing legislation seem to promise an ever-shrinking and more restrictive definition of "fair use purposes" particularly in the digital world. Archivists face the possibility that they may end up providing access to collections in a pay-for-use services digital environment with all funds received going to intellectual property rights holders.

State and federal privacy and publicity laws raise serious concerns over what may be made accessible and how, as do recent publications by culture groups on their wishes to gain legal control over materials already in the public domain or materials created by non-group members.[16] In our litigious times many of these issues are likely to be resolved in the courts or by Congress, rather than by archivists. How can archivists balance the complex and often contradictory requirements and needs of donors, copyright holders, creators, individuals who are documented, scholars, and the general public?

Archival budgets are being seriously eroded by increasing costs, decreased budgets, fewer staff, more users, burgeoning information, increasingly unstable information formats, changing professional information standards and practices, revised laws on fair use and copyright, and institutional restructuring and instability. Simple neglect alone is enough to ensure disaster. The looming dark ages of information loss present us with a renewed mission to save what we can despite our institutional constraints.

What Can We Save?

At the millennium, faced with a new digital dark age of information loss, archivists are re-examining our appraisal strategies. We don't want to be crushed under the weight of the past, nor can we afford to save everything.

Yet, we are aware as never before that our records must reflect the full diversity and complexity of our world, rather than becoming an edited compendium that celebrates a specific world view or a single group. Real archives, like the human unconscious and memory, contain some materials that will be unpopular. A real archives is a by-product of the full range of human actions, rather than the neatly edited version of reality presented by most publications.

Who decides what is preserved and what is destroyed? How do we ensure that the record of "what did he know and when did he know it" is not lost? How do we ensure the memory of the holocaust, slavery, women's suffrage, and Native American disenfranchisement? How do we record the human mistakes, the average day, as well as our best and brightest moments, so that we can learn as well as celebrate? How do we ensure that what we save is authentic, of enduring value, and accessible? Some options are described below.

Working with records creators and users and teaching these individuals to:

- save digital master files in common non-proprietary file formats like TIFF

- avoid using compression when creating master digital files to facilitate future use; instead use compression on derivative or copy files as necessary or desirable
- store master files offline in a redundant array of independent disks, which use multiple servers to back-up data in several different servers simultaneously
- maintain software and hardware and move files to the next generation of software as necessary to keep them useful
- incorporate standard color bars and measuring scales within a digital file to aid in management and viewing
- keep a systematic record of file modifications and changes
- capture and manage file metadata (documentation about digital files) to help when accessing, managing, or viewing the files
- exercise or use all data regularly to ensure full functionality

Some archivists believe we can afford to wait for a technological solution that will not require any change in our relationship with records creators. If technology doesn't provide this answer, these archives will be dependent on whatever knowledge is captured and maintained by the various affiliated user communities in whatever format the communities use. Archives might support these creator or user communities by providing guidance on how to create permanent and durable records and how to manage and preserve their content and their software and hardware over time.

Intervene and statistically sample the digital realm, avoiding the use of human judgment in selection. Work under the assumption that we can't count on a technological fix. A sample has the merit of providing a non-editorial sketch of the whole, although much material of proven and enduring value will be missed simply because it doesn't fit the sample profile. Brewster Kahle's Internet Archives, a digital backup of the Internet taken at regular intervals, captures this data. The question remains: who can afford to continue to manage and migrate this information and provide access to it over time? Kahle recently gave a 12 terabyte copy of his Internet Archives to the Library of Congress, which must determine whether it can afford the management mortgage on this gift.

In ancient India, the gift of a white elephant often bankrupted the state treasury while conferring high status; the Internet Archives may be a similar gift with much status, high user demand, and a punitively high cost to manage through time.

Intervene and select items based upon traditional archival selection criteria such as institutional mission, audience, value in relationship to the mission, usage level of the items, and risk of loss. These criteria have the merit of being based upon what our users want and what history tells us is valuable. However, the resulting archives will reflect a particular worldview rather than the full range of human experience. The advantage of this approach is that it reflects seasoned human judgement and it may be scaled to suit the repository's budget since the repository is selecting in priority order.

Hybrid approach: use the best of all three methods above. First, work with the records creating and using communities to encourage them to create long-lived and durable records, manage them effectively over time, and responsibly provide the records to archives when the creators or the users are done exercising the data. Preserve and manage the hardware and software necessary to make these user-selected materials accessible. Next, statistically sample everything, to an extent that is economically feasible by the archives. Third, select any items that fit the collecting statement of the archives and which are evaluated as having high value to the archives and its audience. Allow value, use, and risk factors to influence the selection process. Finally, pray for technological help and funding from the government, individuals, and the organizations benefiting most from the new technologies, the software and hardware firms.

New Partnerships to Approach

Even as archives face some of the greatest challenges since the first archivists assembled collections of clay tablets in ancient Assyria, we are finding some powerful new partners, including:

The Edutainment Community. Entertainment when wedded with education equals "edutainment." This is best illustrated by the huge growth in specialized book clubs, cable television channels, special interest groups, and web sites focusing on culture, history, natural history, or similar issues, such as:

The History Channel
The Learning Channel
The Knowledge Channel
The Discovery Channel

New Internal Uses, such as:

Geographic Information System demands for historic maps and plans, which result in massive databases that allow land-based agencies, states, governments, and historians to study and know areas in ways previously unimaginable

Government Performance and Results Act (GPRA), which uses baseline data, available often only in archives, to track progress toward accomplishments

Electronic Freedom of Information Act (E-FOIA) mounting of frequently requested federal and state documents on the Web to facilitate access to federal records

Affiliated User Communities, such as:
Civil War re-enactors
collectors
hobbyist networks, such as railroad buffs
clubs, such as historic preservation aficionados

Information Brokers, such as:
Contract researchers, who conduct research for a fee
Picture researchers, who find imagery for films, books, and articles
Online Fulfillment Services, such as Image Directory and Corbis Media, who post the images for others to aid in their resale
Online Rights Management Services, who help organizations track and manage intellectual property rights

Foundations, Organizations, Universities, and Consortia, such as:
The Council for the Preservation of the Anthropological Record's work on preserving the papers of anthropologists (see *CRM* Vol. 18, No. 9, p. 34)
Council on Library and Information Resources focus on preserving the digital record
The Long Now Foundation
Brewster Kahle's Internet Archives
The Getty Information Institute's "Time and Bits" Conference
Northeast Document Conservation Center's "School for Scanning"
Universities, such as:
Cornell University
Harvard University
the University of California at Berkeley
the University of Pittsburgh
Yale University

No organization or profession working alone can preserve our knowledge and historical evidence, or ensure the survival of our information and make it accessible to the insatiable audiences who demand it. We must work together as allied professions and organizations to share our expertise and resources if we are to ensure the survival of our data, information, evidence, and knowledge for future generations. We must explore all options open to us with open minds that are eager to share the message of what is at risk. This legacy, which safely stores our factual observations for future theorists and managers, our information for later adaptive re-use, and our professional knowledge and evidence for enhancement of our organizations and professions, is our greatest gift to the future.

Notes

1 Carlin, John W. "Records Everywhere, But How Are They Going to Survive?" in *The Record,* 5:1, September 1998, p.3.

2 Kahle, Brewster. "Setting the Stage: Summary of the Initial Discussion," in *Time & Bits: Managing Digital Continuity.* Edited by Margaret MacLean and Ben H. Davis, Santa Monica, CA: The J. Paul Getty, Trust, 1998, p. 39.

3 Ibid.

4 Hillis, Danny. "Public Session: Panel Discussion," in *Time & Bits: Managing Digital Continuity.* Op cit., p.42.

5 Ibid, pp. 18-20.

6 Marcum, Deanna. *CLIR Issues,* number 2, March/April 1998, pages 1-3.

7 Hillis, Danny. "Public Session: Panel Discussion," in *Time & Bits: Managing Digital Continuity.* Op cit., p.42.

8 Lyman, Peter in *Time & Bits: Managing Digital Continuity.* Op cit., p.11-12.

9 Carlin, John W. "Records Everywhere, But How Are They Going to Survive?" in *The Record,* 5:1, September 1998, pp 1-3.

10 Taylor Shelley Z. "Ongoing Digital Dialogue: The Time and Bits Threaded Discussion," in *Time & Bits: Managing Digital Continuity.* Op cit., p 65.

11 Carlin, John W. "Records Everywhere, But How Are They Going to Survive?" in *The Record,* 5:1, September 1998, pp 1-3.

12 Eiteljorg, Harrison. "Archiving Archeological Data in the Next Millennium" *CRM,* 21:6 (1998), pp. 21.

13 Puglia, Steven. "Cost Benefit Analysis for Reformatting Option," Speech at AfterImages: Reformatting Visual Materials in a Digital World, Conference of NEDCC and NPS at NARA Archives II, September 17-19, 1998.

14 Public Citizen, Inc., et. al, v. John Carlin, Archivist of the United States, U.S. District Court, 1997 (Civil Action 96-2840 (PLF)).

15 Kiesling, Kris "What Is the Encoded Archival Description Standard?" in this issue of *CRM,* page 28.

16 Brown, Michael F. "Can Culture Be Copyrighted?" in Current Anthropology, 19:2, April 1998, pp.193-206 and "Cultural Records in Question: Information and its Moral Dilemmas," in *CRM* 21:6, (1998) pp. 15-17.

Diane Vogt-O'Connor is Senior Archivist, Museum Management Program, National Park Service.